NCAR

DTC Developmental Testbed Center

Evaluation of New Analyses and Methods for Verification of Cloud Predictions

Barbara Brown, Tara Jensen, John Halley Gotway, Kathryn Newman, Eric Gilleland, Tressa Fowler, and Randy Bullock

Testbed and Proving Ground Workshop
April 25, 2017

National Center for Atmospheric Research

Motivation and Goals

NCAR

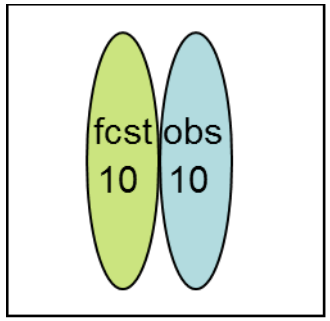
- Motivation
  - Clouds have important impacts on activities of the US Air Force and are a prime focus of the 557th Weather Wing
  - Skill of cloud forecasts impacts decision making (e.g., uncertainty in cloud cover predictions can change operational decisions)
- Goals
  - *Long-term*: Create a meaningful cloud verification "index"
  - *Short-term*: Identify useful components of such an index

Approach

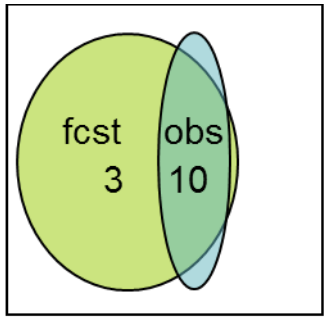
1. **Standard methods** based on traditional metrics (continuous, categorical)
2. Investigate **object-based and distance metrics** to provide forecast quality information that
   - Provides diagnostic, user-relevant information
   - Includes methods not subject to "hazards" of traditional verification (e.g., "Double Penalty")

Initial focus on CONUS, fractional coverage (TCA = Total Cloud Amount)

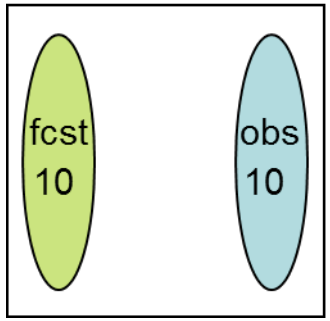
Secondary: Global forecasts, ARM observations



Hi res forecast
RMS ~ 4.7
POD=0, FAR=1
TS=0



Low res forecast
RMS ~ 2.7
POD~1, FAR~0.7
TS~0.3

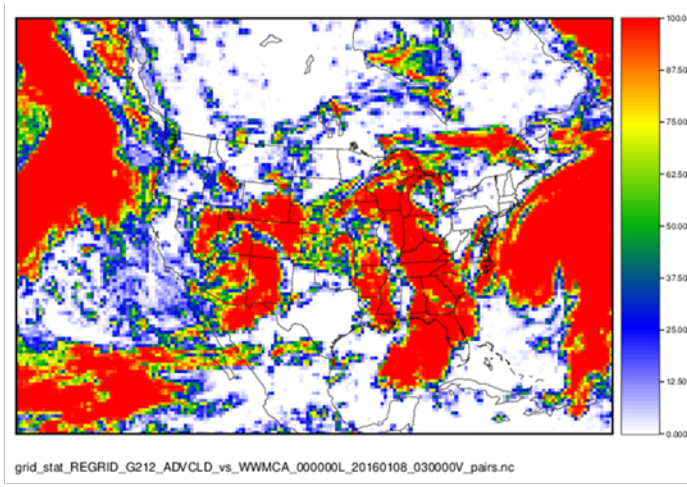


Observations, Analyses, and Forecasts

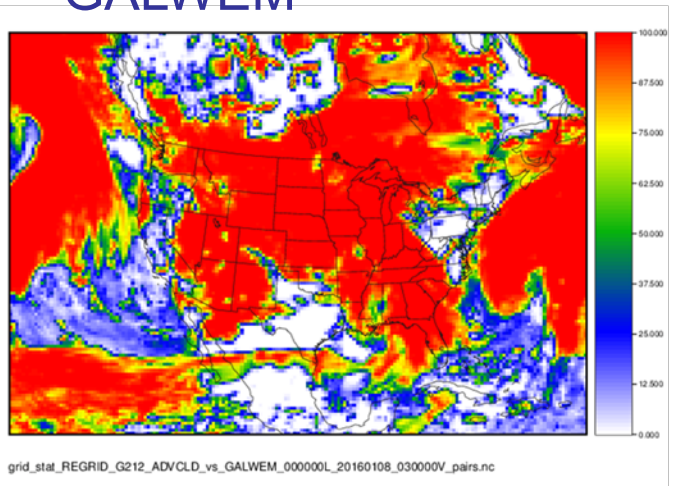
NCAR

- **"Observations" and Analyses**
  - METARs (but not shown here)
  - WWMCA (gridded World-Wide Merged Cloud Analysis)
  - WWMCA-R (WWMCA updated in post-analysis with all obs available)
  - ARM site Total Cloud Amount (TCA) for 4 locations
- **Forecasts**
  - 2 global models (72 h)
    - **GALWEM** (AF implementation of UK Unified Model)
    - **GFS** (NCEP Global Forecast System)
  - **DCF** (Diagnostic Cloud Forecast)
    - Bias-corrected GALWEM and GFS
  - **ADVCLD**: Advection (persistence) model (9 h)
- Sample data for 4 seasons (1 week each)
- NCEP grid 212 (polar stereographic; 40 km)
- Model Evaluation Tools (MET) and Spatial-Vx R package used for all analyses

WWMCA



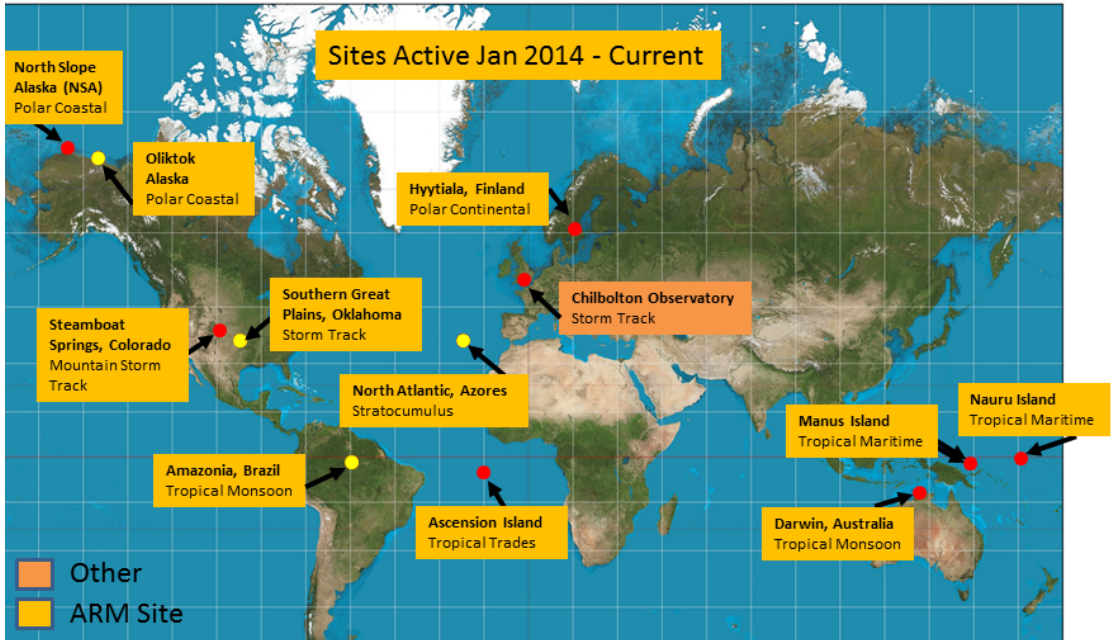
GALWEM



Conclusions First…

- **Continuous methods** (RMSE, MAE, etc.) do not provide much useful information regarding TCA performance – primarily due to discontinuous nature of clouds
  - Edges
  - Tendency of products toward 0 or 100% values
- **Point observations are less useful overall than satellite-based analyses due to limited availability globally**
- **Categorical methods** (POD, FAR, etc.) are more useful for answering relevant questions about cloud occurrence
  - Especially when presented in a diagnostic multivariate form
- **Object-based methods** have promise of providing useful information – when configured appropriately
- **Distance metrics** can provide interesting diagnostic information – but need to be explored more

ARM sites

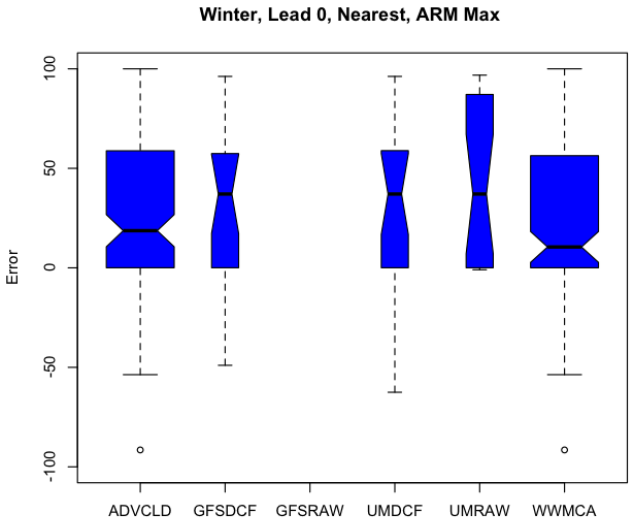


Cloud amount measurements from Total Sky Imager used for the evaluations

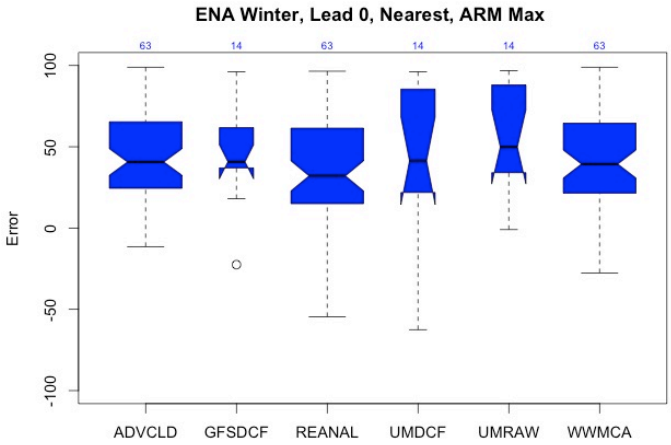
Limited ARM data obtained for 4 ARM sites during our 4 periods

- Oliktok, Alaska (OLI)
- Southern Great Plains, Oklahoma (SGP)
- North Atlantic, Azores (ENA)
- Amazonias, Brazil (MAO)

Example ARM comparisons



4 ARM stations
Winter
Lead=0



ENA (Azores)
Winter
Lead=0

Errors vary by

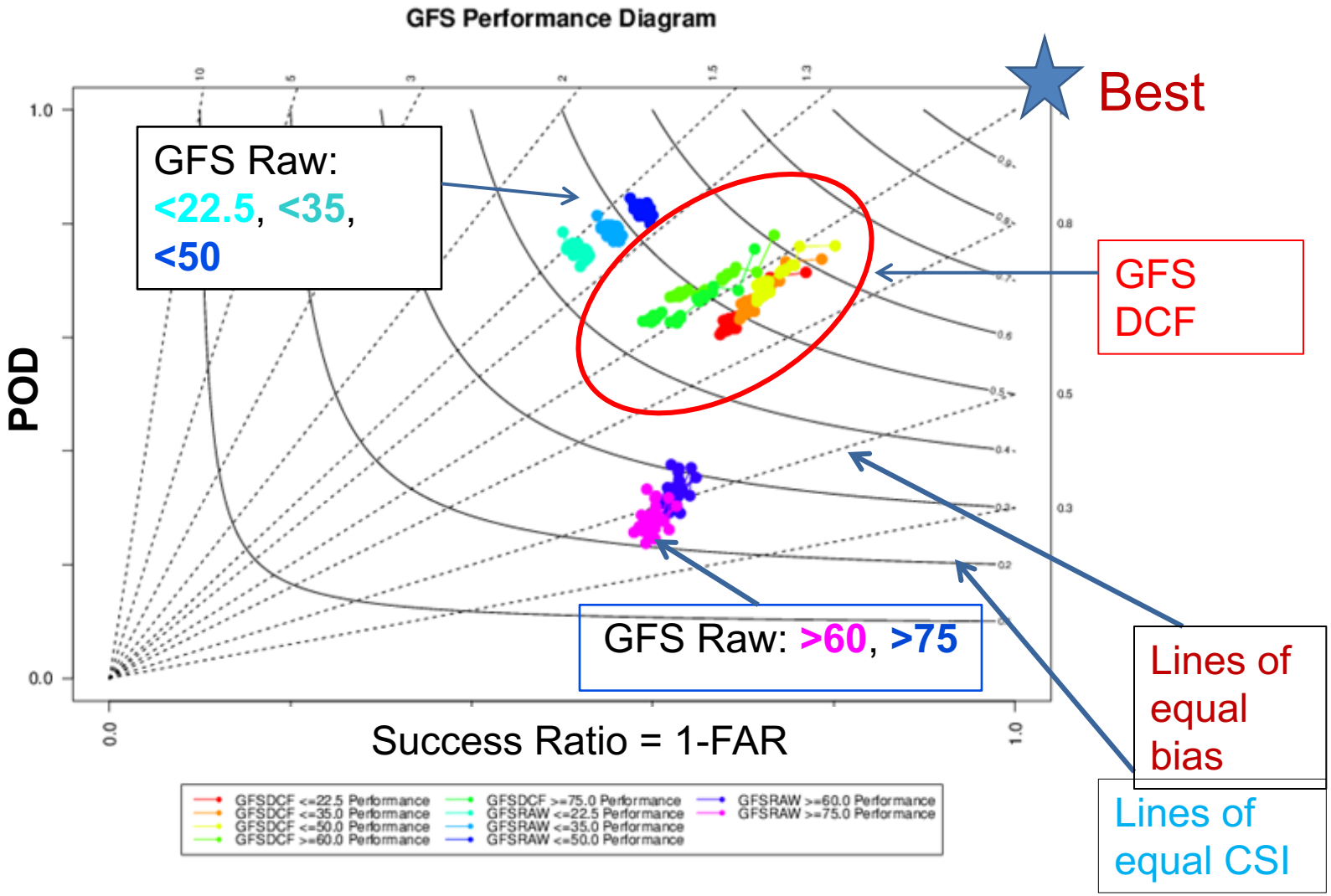
- Location
- Matching approach (e.g., Max vs. Average)
- Neighborhood size (e.g., nearest, 9, 16 gridpoints, etc.)

Although active sensors should be the “best” datasets for comparison, limitations in data availability limit their potential usefulness

- Difficult to combine results across locations

Gridded comparisons: Categorical statistics

Performance Diagrams using WWMCA-R as the verification grid



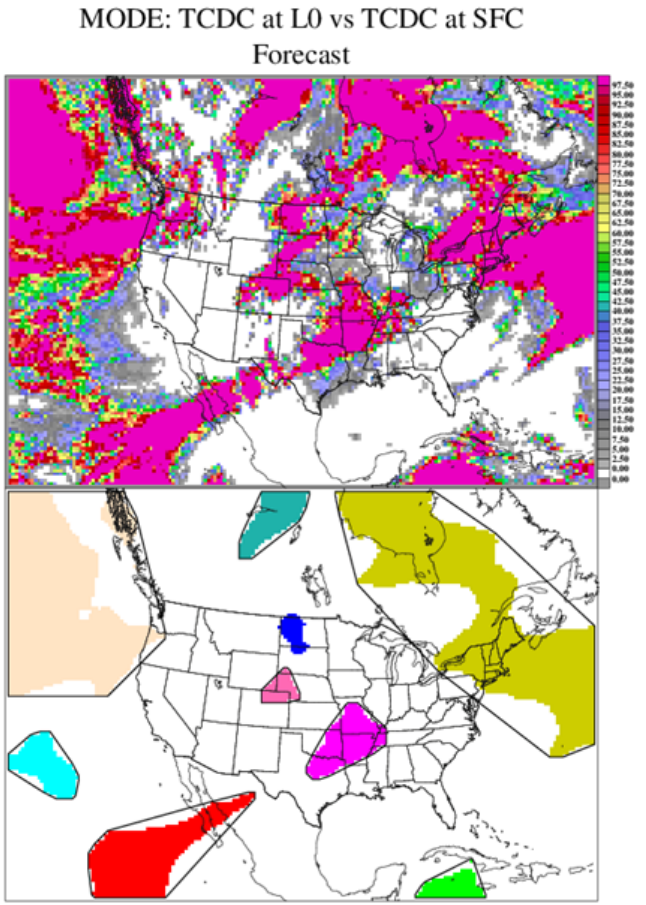
After Roebber (2009)

MODE Object-Based Approach

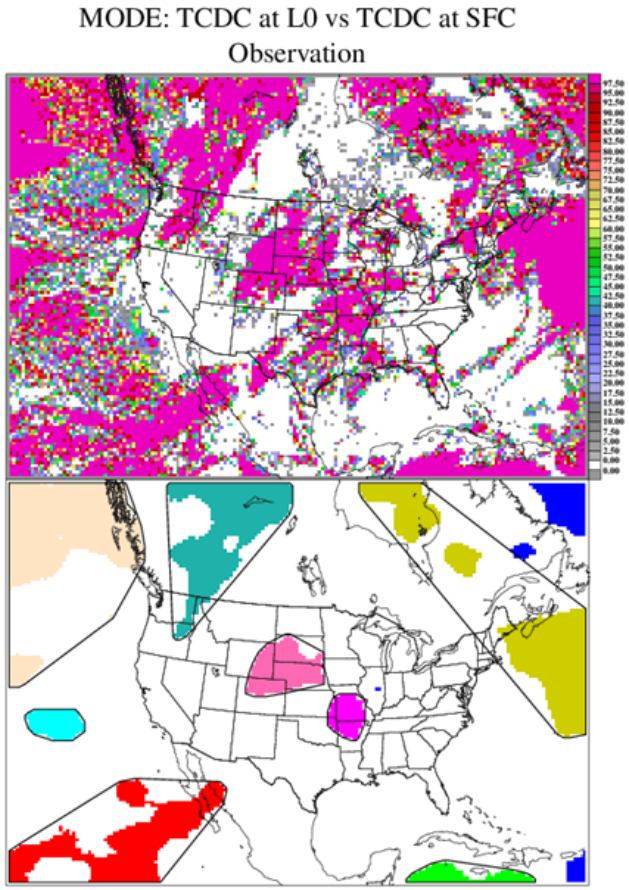
MODE in a nutshell:

- Identify objects
- Measure and compare object "attributes" (e.g., size, location, intensity)

GALWEM

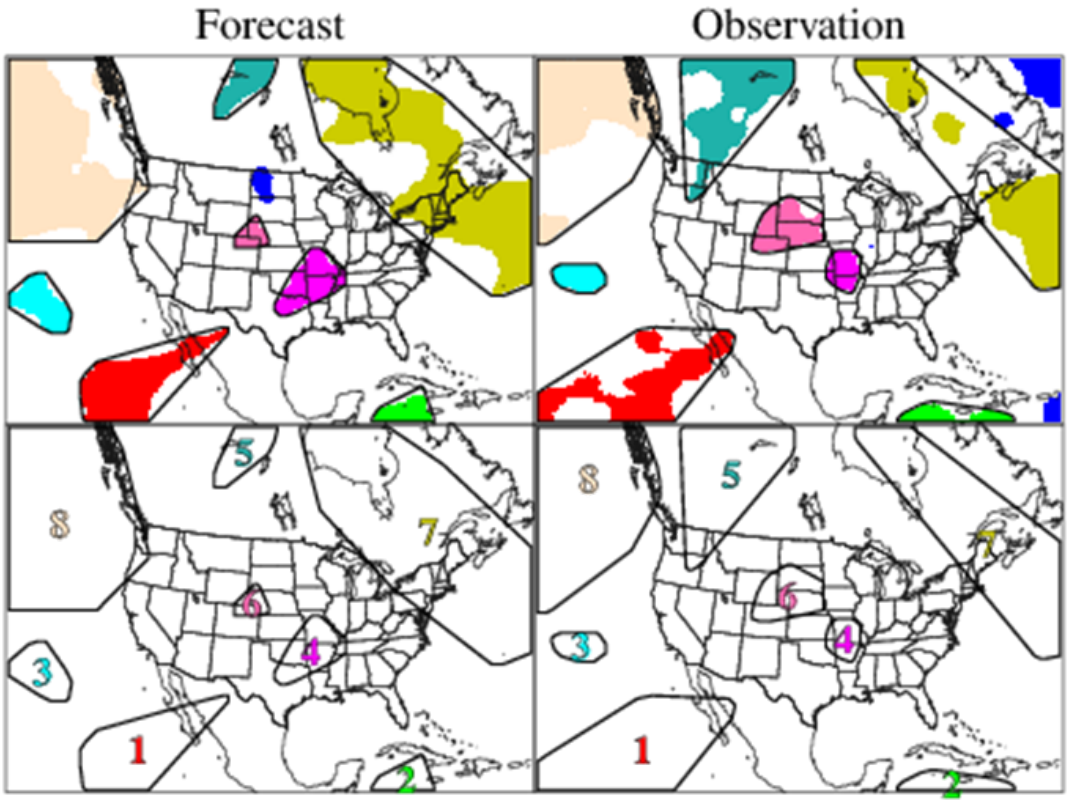


WWMCA



11 November 2015
Cloudy Threshold (TCA > 75)

Cluster Object Information

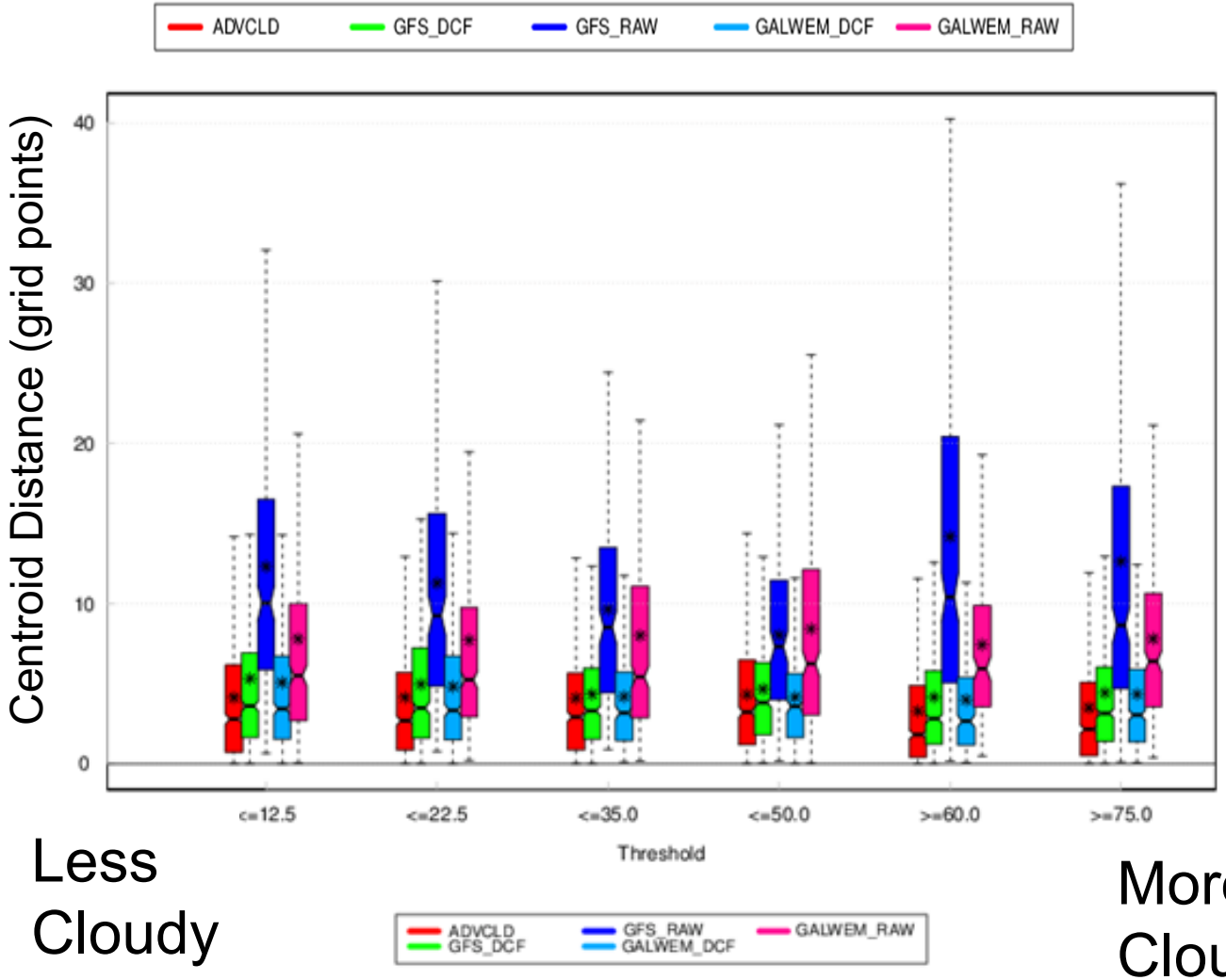


- Some displacement of all clusters
- Large area differences, for some objects

… Etc.

CLUS PAIR	CEN DIST	ANG DIFF	FCST AREA	OBS AREA	INTER AREA	UNION AREA	SYMM DIFF	FCST INT 50	OBS INT 50	FCST INT 90	OBS INT 90	TOT INTR
1	8.53	10.08	689	816	504	1001	497	100.00	100.00	100.00	100.00	1.0000
2	6.18	10.69	131	138	87	182	95	100.00	100.00	100.00	100.00	1.0000
3	9.80	35.64	247	145	33	359	326	89.00	100.00	100.00	100.00	0.9411
4	4.69	51.94	299	130	121	308	187	100.00	100.00	100.00	100.00	0.9158
5	16.56	13.02	229	829	196	862	666	100.00	100.00	100.00	100.00	0.9018
6	3.47	19.33	81	305	81	305	224	100.00	100.00	100.00	100.00	0.8958
7	11.74	2.27	2366	1049	1001	2414	1413	100.00	100.00	100.00	100.00	0.9407
8	15.77	38.71	1921	1157	773	2305	1532	100.00	100.00	100.00	100.00	0.9607

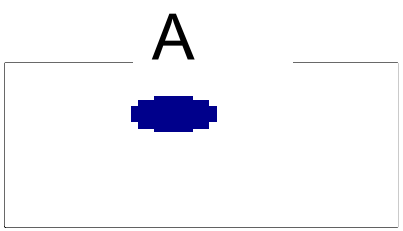
Example MODE summary result: Centroid Distance

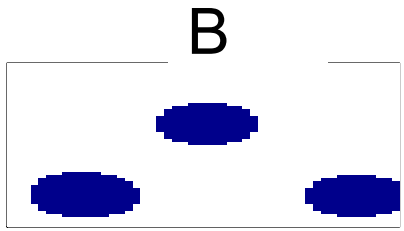


Less Cloudy

More Cloudy

Distance Map Measures

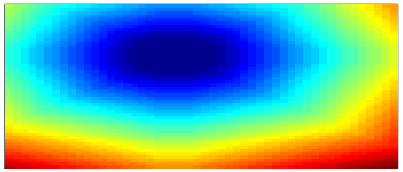


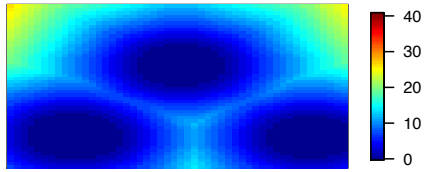


Distance map of A

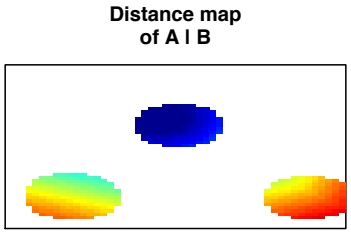
Distance map of B

d(x, A)

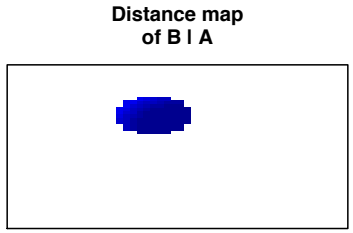




d(x, B)



d(x, A | B)

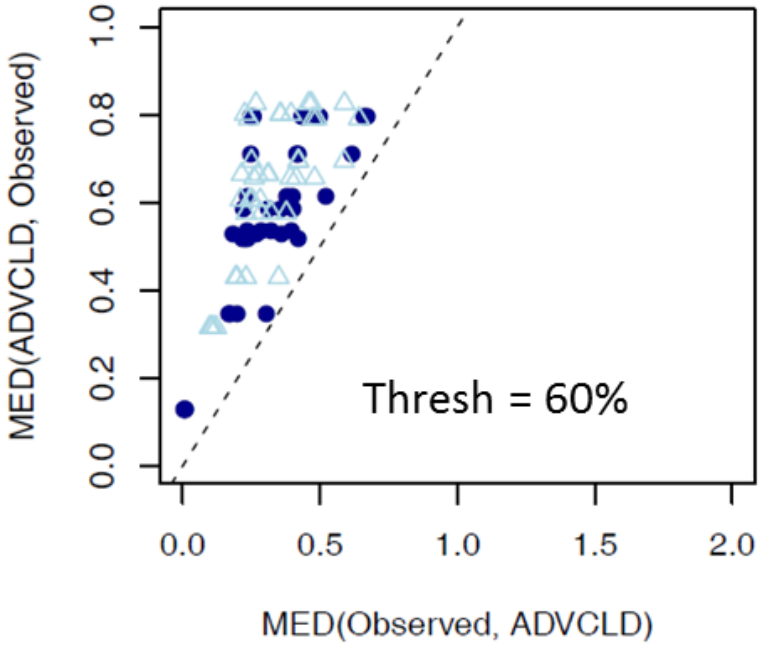


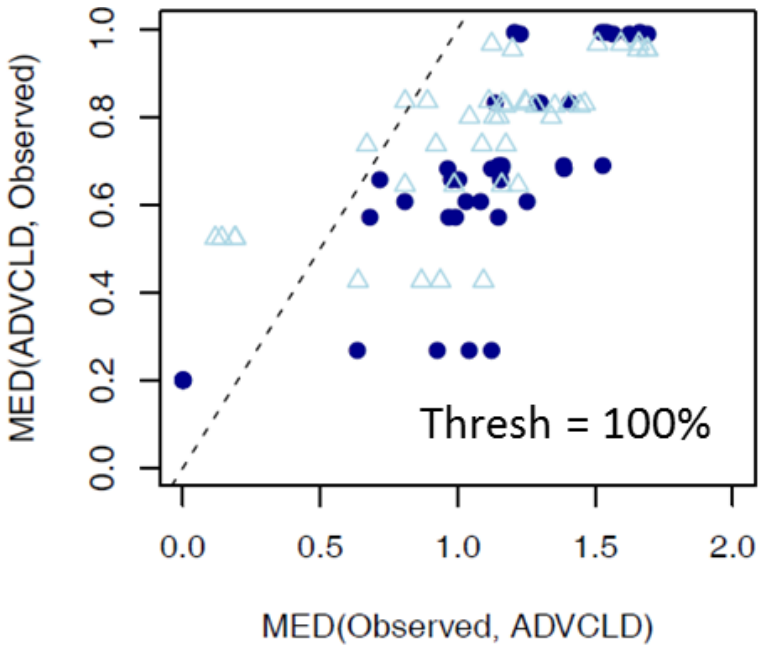
d(x, B | A)

MED(A, B) = $\Sigma_{\mathbf{s}\ \text{in}\ D}\ d(x, A \mid B) / N$,

s are locations in the domain, D, and N the total number of grid cells.

Mean Error Distance





Examine average error distance from all forecast points to the nearest obs point [**MED(forecast, obs)**], and from all obs points to the nearest forecast point [**MED(obs, forecast)**]

- Above diagonal: Misses
- Below diagonal: False alarms

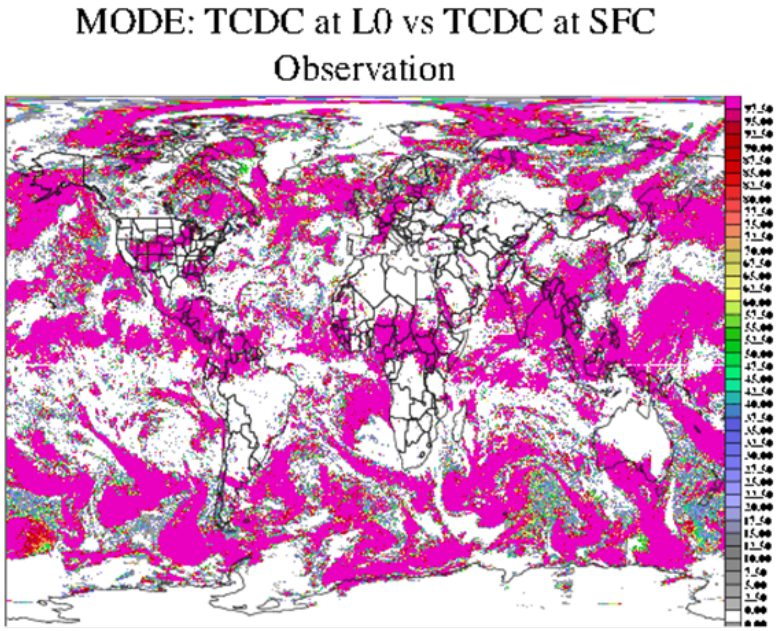
Gilleland 2017 (WAF)

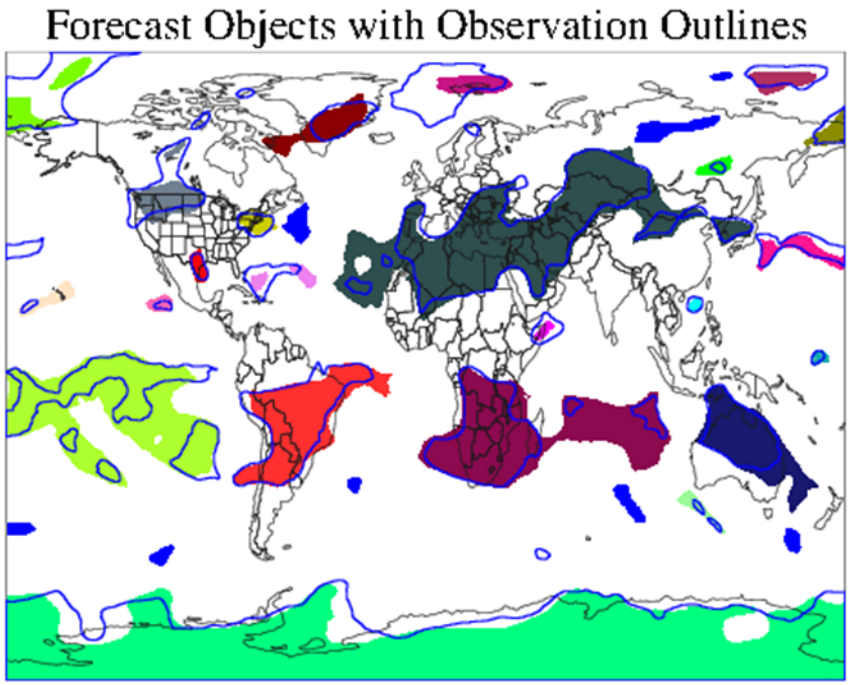
Other promising approaches:

- **Hausdorff and Baddeley Delta metrics**
- **Image warping**
- **Geometric measures**

Conclusions

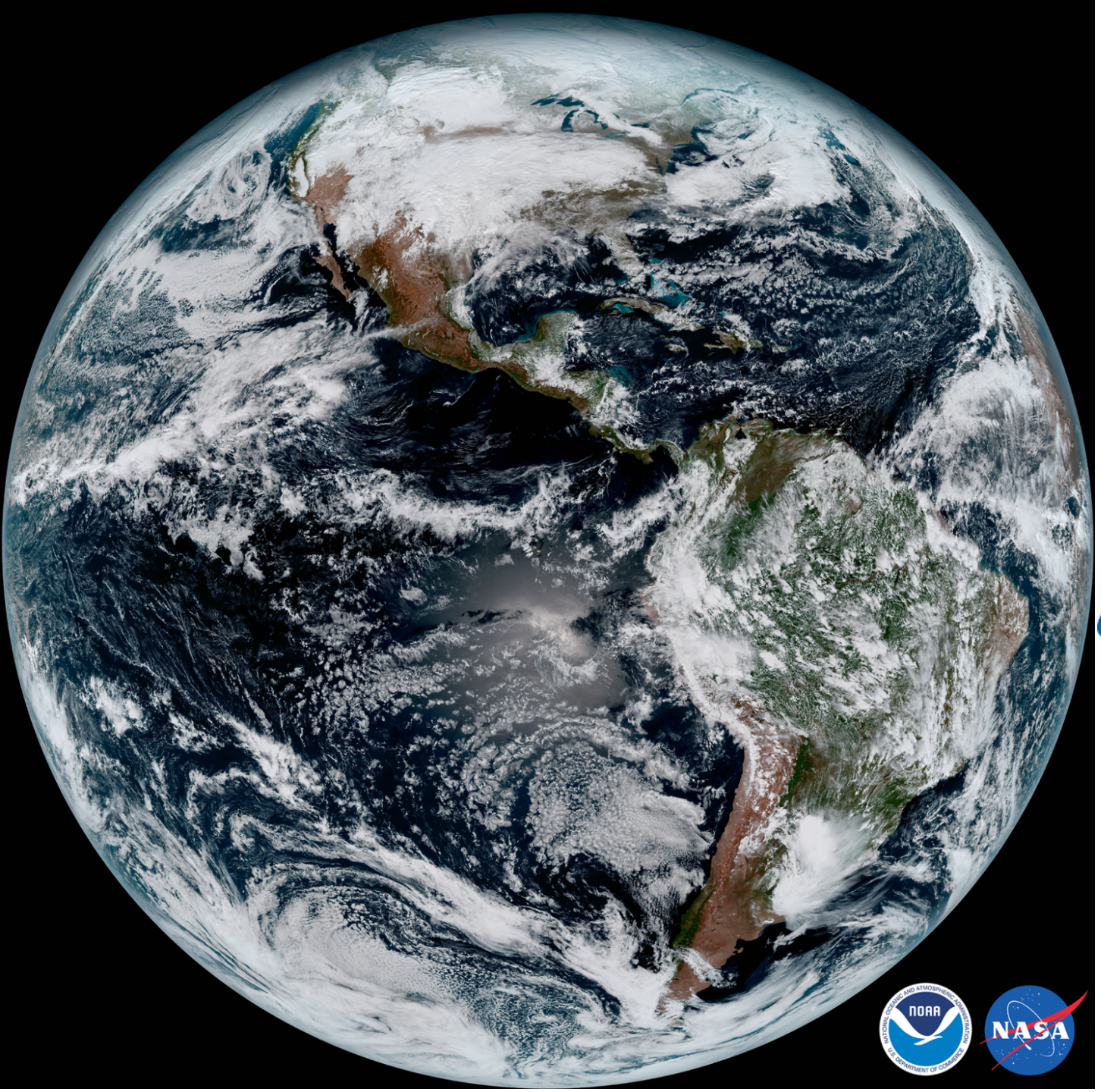
- Categorical methods are the most useful "traditional" approach for evaluating TCA
  - Diagnostic plots (box plots, performance diagrams aid in interpretation of results)
- Spatial and distance metrics have a lot of benefits and are promising approaches
- On a global scale, MODE is especially useful for *evaluation of non-cloudy areas*





Future Work

- Further *tests of distance and geometric methods*, and other *spatial approaches*
- Evaluation methods for *bases, tops, layers, and other user-relevant variables*
- Use of additional *active sensors*
- Take into account some aspects of *observational uncertainty* – e.g., pixel age



Thank You

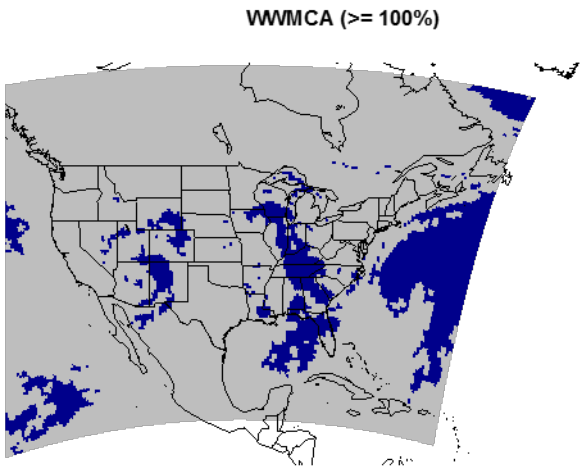
DTC Developmental Testbed Center

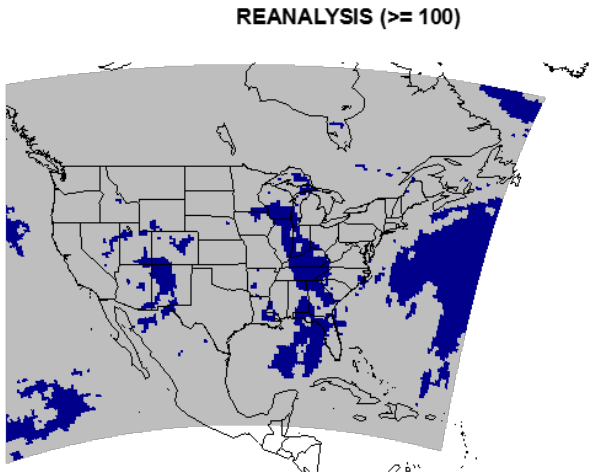
NCAR

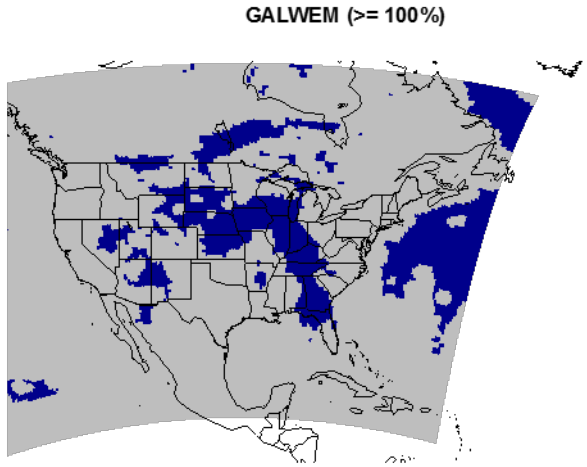
Geometric Approach

NCAR

Measure and compare Geometric characteristics of objects/areas: Connectivity (C), Shape (S), Area (A) (AghaKouchak et al. 2010; *J. Hydromet*)







	WWMCA-R	WWMCA	GALWEM
C_{index}	0.674	0.713	0.801
S_{index}	0.398	0.408	0.436
A_{index}	0.173	0.180	0.216